3D Reflector Localisation and Room Geometry Estimation using a Spherical Microphone Array

Michael Lovedee-Turner[1, a)] and Damian Murphy[1]
AudioLab, Communication Technologies Research Group, Department of Electronic Engineering, University of York, UK

The analysis of room impulse responses to localise reflecting surfaces and estimate room geometry is applicable in numerous aspects of acoustics, including source localisation, acoustic simulation, spatial audio, audio forensics, and room acoustic treatment. Geometry inference is an acoustic analysis problem where information about reflections extracted from impulse responses are used to localise reflective boundaries present in an environment, and thus estimate the geometry of the room. This problem however becomes more complex when considering non-convex rooms, as room shape can not be constrained to a subset of possible convex polygons. This paper presents a geometry inference method for localising reflective boundaries and inferring the room's geometry for convex and non-convex room shapes. The method is tested using simulated room impulse responses for seven scenarios, and real-world room impulse responses measured in a cuboid-shaped room, using a spherical microphone array containing multiple spatially distributed channels capable of capturing both time- and direction-of-arrival. Results show that the general shape of the rooms is inferred for each case, with a higher degree of accuracy for convex shaped rooms. However, inaccuracies generally arise as a result of the complexity of the room being inferred, or inaccurate estimation of time- and direction-of-arrival of reflections.
 [http://dx.doi.org(DOI number)]

[XYZ] Pages: 1–13

I. INTRODUCTION

A room impulse response (RIR) is the characteristic response of a room to excitation from a known broadband test signal. These RIRs are comprised of a superposition of the direct source-to-receiver component, discrete early reflections produced through limited interactions with boundaries within the room, and a densely-distributed and exponentially decaying reverberant field. RIRs are therefore representations of the reverberant characteristics of a room, and are uniquely defined by the geometric constraints of the room and the source/receiver locations. This property makes RIRs an invaluable resource for acoustic analysis and acoustic scene rendering. One key application for RIR analysis is geometry inference - a form of acoustic analysis focusing on the estimation of a room's geometry from the reflections captured across a number RIRs. The ability to create an accurate model of the geometric constraints of a given room has numerous further applications in acoustics ranging from: sound source localisation, room acoustic simulation, spatial audio, audio forensics, and the acoustic treatment of rooms. Geometry inference algorithms can typically be split into one of two categories[1]: image-source reversion[1–5] and direct localisation[1,6–10].

Previous work has been shown to be able to infer the geometry of a room from a set of RIRs, however, they are only presented for the case of a simple cuboid room, and in some cases require large numbers of RIRs to work successfully. Furthermore, most do not present error management techniques for reducing the impact that inaccurately detected reflections or incorrectly computed reflection paths have on the inferred geometry.

In this paper the problem of geometry inference is considered from the perspective of both convex and non-convex 3D room shapes, which has not been considered previously in the literature. To achieve this the Eigenbeam Detection and Evaluation of Simultaneously-Arriving Reflections (EDESAR) reflection detection method is proposed based on spherical harmonic decomposition of spatial room impulse responses (SRIRs), which allows simultaneously arriving reflections from different direction-of-arrivals (DoA) to be detected as individual reflections. Furthermore, an extension to existing image-source reversion techniques is proposed, the Acoustic Reflection Cartographer method. A geometry validation process is then proposed to refine the inferred room's geometry to ideally that of the desired room. The following assumptions are made when inferring the room's geometry,

- Source-receiver distance and room temperature are known *a priori*
- There is at least a 50 cm distance from the source and receiver to the boundaries (half the standard

[a)] mjlt500@york.ac.uk; Funding was provided by a UK Engineering and Physical Sciences Research Council (EPSRC) Doctoral Training Award.
Code will be made available at: 10.5281/zenodo.2563643

measurement distance[17] allowing for analysis of smaller/complex rooms)

- Reflections have a dominant specular component
- Boundaries define a closed room
- Floor and ceiling are parallel to each other
- Walls are perpendicular to both the floor and ceiling

The paper is organised as follows: section II presents background literature in the field of geometry inference, section III presents the problem formulation and proposed methodology, section IV discusses the testing methodology employed to assess the proposed methods accuracy, section V and VI will present and discuss the results, and section VII concludes the paper.

II. BACKGROUND

Image-source reversion defines the set of techniques that first estimate the locations of image-sources from the ToA, and in some cases DoA, of reflections arriving at one or more receivers[1].

Dokmanic *et al.*[2] proposed a technique exploiting the properties of Euclidean distance matrices to search for reflections with a common image-source across RIRs measured at five receiver locations, each positioned such that they receive a first-order reflection from each wall. The image-source locations are then defined as the common point of intersection across spheres centred around the receiver positions, with radius values derived from the ToA for each reflection. This method required *a priori* knowledge of the source and receiver position, and assumed that the room was convex in shape.

Arteaga *et al.*[3] used the source-receiver distance and reverberation time computed from a measured RIR to define the geometries for a set of possible cuboid rooms. The RIR for these possible rooms are simulated using the image-source method, and a goodness-of-fit is computed between the simulated and measured RIRs to find the cuboid room that produces a RIR similar to that measured.

Ribeiro *et al.*[4] used a least-squares minimisation technique to fit synthetically generated reflections to a measured RIR. This process detects a sparse set of reflections with approximately known DoA. Image-sources are then generated from the ToA and DoA, the boundary locations are then estimated from the image-source positions. These boundaries are only considered valid if at least one second- or third-order reflection is also detected for the boundary. This technique assumed a convex-shaped room when inferring the rooms shape, and required *a priori* knowledge of the source and receiver positions, and the microphone arrays response to signals from a grid different DoA.

Tervo and Tossavainen[5] employed a maximum likelihood approach to find the location of the image-source that maximised a utility function from a randomly generated set of points in space. The fist-order reflections were used to define the locations of the boundaries. If an already defined plane can be used to define the location of an image-source it is assumed to be a higher-order reflection and is ignored. This approach assumed that the room is convex in shape, and that each RIR contains a discrete detectable first-order reflection from each boundary.

Remaggi *et al.*[1] used an adaptation of the Dynamic Programming Projected Phase-Slope Algorithm[11] to detect reflections in a RIR, and a delay-and-sum beamformer to find the DoA. The image-source locations are then defined in the spherical coordinate domain using the azimuth, elevation, and ToA computed for each reflection. Remaggi *et al.* assumed that the source and all image-sources were at least 2.1 m away from the microphone array, and that at least four measurement positions existed - although more were used in testing.

Alternatively, direct localisation techniques aim to infer boundary positions directly without extracting further information about the reflection paths[1].

Remaggi *et al.*[1] and Nastasia *et al.*[7], used ToA to define ellipsoids with foci on the source and receiver locations, which are known *a priori*. Through RIR measurements at different source and/or receiver locations, a boundary is defined as being at the point where there is a common tangent across ellipsoids produced by a reflection common across all RIRs. Assumnptions made by Remaggi*et al.* are as above. Nastasia *et al.* assumed a minimum and maximum reflector distance from the origin to remove unwanted planes.

Kuster *et al.*[6] proposed a technique based on inverse-wave field extrapolation of reflections captured using a linear-array of microphones. The use of large numbers of measurement positions allowed the approach to produce a detailed mapping of a single boundary in an environment.

Filos *et al.*[8] proposed a method for estimating room geometry using a seven microphone array split into three sub-arrays of five microphones. Each sub-array is positioned to localise planes on two axes, xy, xz, and yz, and RIRs are captured using a source position for each boundary in the room. The Hough transform is then used to find the line that is tangential to each ellipse defined by the ToA for the first arriving reflection in each RIR set. Intersections between the infered lines are then used to define the room's geometry[8]. However, the array type used limits this approach to geometry inference to the case of convex rooms.

Zamaninezhad *et al.*[9] estimated the distances between two reflective boundaries from the location of the main resonant frequencies within the frequency domain representation of the RIR, the room transfer function. The main resonant frequency is detected through minimisation of a cost function of the possible resonances within the measured room transfer function. To infer the location of the boundaries, it is assumed that one boundary is defined at $x = 0$ which is closer to the source.

Baba *et al.*[10] proposed an extension to the ellipsoid based methods. Firstly, they proposed the use of a linear Radon transform, an image processing algorithm used for line detection, to locate common reflections across a stacked set of RIRs measured across multiple loudspeaker positions. Knowledge of the array geometry and ToA is then used to create an image-microphone. The point of reflection is then located as the point of intersection between the ellipse and a line going from the image-microphone to the source. It is assumed that each wall has a uniform-linear loudspeaker array parallel to it, and therefore the geometry of the room is constrained to the geometry of the loudspeaker array.

These methods require multiple measured RIRs recorded at different points in the room, with precise positioning of each source and receiver so that a discrete and detectable first-order reflection for each boundary exists. This inherently limits the application of these methods to convex rooms, where the source and receiver placement requirements can be met, and allows boundaries that are not detected across the RIRs to be removed. This, however, does not hold for non-convex shaped rooms, where boundaries will not necessarily be visible to all source and receiver positions.

III. PROBLEM FORMULATION AND METHOD

Spherical microphone arrays measure the sound pressure on the surface of a rigid sphere, spatially sampled at the microphone positions distributed across this surface. Given a spherical microphone array, the sound field on the surface of the sphere can be defined using spherical harmonics[12]. In the time-domain, when assuming frequency independent reflection, the RIR $\mathbf{H}(t)$ can be represented as a superposition of Dirac deltas δ steered by the spherical harmonic vector $\mathbf{y}(\Psi_i)$ in the azimuth and elevation DoA, $\Psi = [\theta, \phi]$, of the arriving direct sound and reflections, with amplitude α and time-of-arrival (ToA) τ, with the addition of the time-variant residual noise component $\mathbf{R}(t)$ as,

$$\mathbf{H}(t) = \sum_{i=1}^{\infty} \mathbf{y}(\Psi_i)\alpha_i\delta(t-\tau_i) + \mathbf{R}(t) \tag{1}$$

where the spherical harmonics column vector of order N, $\mathbf{y}(\Psi)$, contains the $(N+1)^2$ spherical harmonics[12],

$$\mathbf{y}(\Psi) = [y_0^0(\Psi), y_1^{-1}(\Psi), y_1^0(\Psi), y_1^1(\Psi), ..., y_N^M(\Psi)]^T \tag{2}$$

where $(.)^T$ denotes transposition and the real valued spherical harmonic of order n and degree m, y_n^m, is computed as[14],

$$y_n^m = \begin{cases} \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(cos\phi)\sqrt{2}cos(m\theta), & \text{if } m > 0 \\ \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(cos\phi), & \text{if } m = 0 \\ \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(cos\phi)\sqrt{2}sin(m\theta) & \text{if } m < 0 \end{cases} \tag{3}$$

where P_n^m is the associated Legendre polynomial of order n and degree m.

Through analysis of the spatiotemporal attributes of the reflections in the SRIR $\mathbf{H}$, the DoA and ToA for individual reflections can be extracted. This information can then be used to define the locale of image-sources that produce the reflections arriving at the receiver. The boundary location can then be estimated from the image-source, and the previous-source that was mirrored in the boundary to produce the image-source, by exploiting the properties of the image-source method. That is, as an image-source is produced by mirroring the previous-source perpendicularly across a boundary, the distance from previous-source-to-boundary and boundary-to-image-source are equal, and the line between the previous-source and image-source is parallel to the boundary's normal. A point on the boundary $\widetilde{\mathbf{b}}$ and the boundary's normal $\widetilde{\mathbf{n}}$ can therefore, from[2], be estimated as,

$$\widetilde{\mathbf{b}} = \frac{\widetilde{\mathbf{s}} + \mathbf{s}}{2} \tag{4}$$

$$\widetilde{\mathbf{n}} = \frac{\widetilde{\mathbf{s}} - \mathbf{s}}{||\widetilde{\mathbf{s}} - \mathbf{s}||} \tag{5}$$

where $\widetilde{\mathbf{b}}$ is a point on the candidate boundary, $\widetilde{\mathbf{n}}$ is the candidate boundary's normal, $\widetilde{\mathbf{s}}$ is the location of an image-source, and $\mathbf{s}$ is the source location.

Therefore, there are two main stages to the adopted approach: the analysis of the spatiotemporal information contained in the SRIRs (subsection III A), and the inference of the room's geometry from this information (subsection III B).

A. Eigenbeam Detection and Evaluation of Simultaneously Arriving Reflections

To extract the spatiotemporal information, a reflection detection and analysis technique using spherical beampatterns (eigenbeams) is proposed, the Eigenbeam Detection and Evaluation of Simultaneously Arriving Reflections (EDESAR) method.

The proposed method analyses the SRIR iteratively over windowed time-frames of 0.45 ms with a 50% frame overlap. These short time-frames allow for reflections arriving from close DoAs but different ToA to be easily detected separately. Before any analysis of the impulse

response is performed, the SRIR is normalised to have a maximum sample magnitude of ±1. To reduce the impact of noise, and reduce the likelihood of false positive detections, time-frames are ignored if they have a maximum sample value across all channels less than a defined threshold value ϵ_a.

The time-frames being analysed are windowed using a Hann window then low-pass filtered at 5 kHz, and high-pass filtered at 100 Hz, reducing the impact of diffuse spectral components as a result of the spatial Nyquist frequency[18], 8 kHz as quoted for the EigenMike EM32[19] as used in this study.

The diffuseness profile for each time-frame is then computed using the Covariance Matrix Eigenvalue Diffuseness Estimation (COMEDIE) algorithm[20] as implemented in the *Spherical-Array-Processing* Toolbox (*'getDiffuseness_CMD'*)[14]. This diffuseness estimator was chosen based on results presented in[20], which showed that the COMEDIE algorithm produced a more robust estimate of diffuseness as a result of being able to disambiguate between multiple correlated/uncorrelated sound sources and spatially diffuse noise[20]. To determine the diffuseness of a time-frame the diffuseness profile is computed for each spherical harmonic order (up to third-order in this study), and if a time-frame has any diffuseness profiles greater than the threshold value ϵ_d it is ignored. This reduces the likelihood of inaccuracies in ToA and DoA estimation, as a result of the number of signals or the signal-to-noise ratio, which would impact the algorithms ability to infer reflection paths, and therefore boundary locations.

To detect directional signals within the filtered time-frame, the Minimum Variance Distortionless Response (MVDR) beamformer[21], based on the *'sphMVDR'* function in the *Spherical-Array-Processing* toolbox[14], is used. The MVDR beamformer was chosen for its ability to minimise the impact of signal variance on the steered response of the microphone array, which therefore produces more accurate predictions of DoA[15]. The MVDR beamformer is computed by steering the response of the microphone array in the spherical harmonic domain, by spatially filtering the response with a weighted spherical harmonics vector. The spherical harmonics vector is weighted to ideally reduce the impact of unwanted noise on the DoA estimation, and is computed from the time-frame's covariance matrix as[16],

$$\mathbf{w}(\Psi) = \frac{\mathbf{R_H}(t-\tau_\omega)^{-1}\mathbf{y}(\Psi)}{\mathbf{y}(\Psi)^T\mathbf{R_H}(t-\tau_\omega)^{-1}\mathbf{y}(\Psi)} \tag{6}$$

where $(.)^{-1}$ denotes matrix inversion, $\mathbf{y}(\Psi)$ is the $[16\times 1]$ spherical harmonic vector computed using the *getSH* function in the *Spherical Harmonic Transform Library*[13], and $\mathbf{R_H}(t-\tau_\omega)$ is the $[16\times16]$ covariance matrix for time-frame τ_ω in RIR $\mathbf{H}(t)$ computed as[14],

$$\mathbf{R_H}(t-\tau_\omega) = \mathbf{H}(t-\tau_\omega)^T\mathbf{H}(t-\tau_\omega) + \frac{\mathbf{I}_{(N+1)^2}}{4\pi} \tag{7}$$

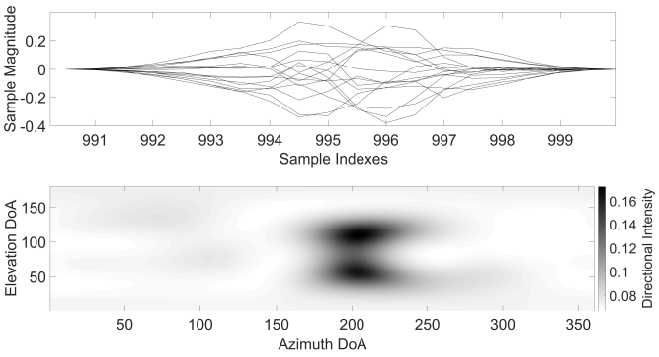

FIG. 1. Top: A typical time-frame for a room impulse response containing two simultaneously arriving reflections, where each line represents a different channel in the third-order spherical harmonic signal. Bottom: The directional spectrum computed for the time-frame, where the darker regions indicate the arrival of strong directional components in the signal.

where $\mathbf{I}_{(N+1)^2}$ is the $[(N+1)^2\times(N+1)^2]$ identity matrix. The MVDR beamformer output is then computed as,

$$\mathbf{\Lambda}(\Psi) = \mathbf{w}(\Psi)^T\mathbf{R_H}(t-\tau_\omega)\mathbf{w}(\Psi) \tag{8}$$

where $\mathbf{\Lambda}(\Psi)$ is the intensity of the signal in the direction Ψ calculated by steering the arrays response with the $[16\times1]$ weighted spherical harmonics vector $\mathbf{w}(\Psi)$.

The directional spectrum for each time-frame is then computed as the directional intensity of the signal steered across a grid of azimuth and elevation positions from $0^\circ \leq \theta \leq 359^\circ$ and $0^\circ \leq \phi \leq 180^\circ$ in one degree increments. An example of this directional spectrum can be seen in Figure 1.

The next step is to detect the darker regions within the directional spectrum, which represent the arrival of directional signals in the time-frame. When considering the nature of the data being analysed as seen in Figure 1, it is evident that two steps are required, separating out spatial regions that overlap and detecting the directional regions in the spectrum. These steps can be easily achieved through use of existing image processing techniques with small modifications to the intensity matrix.

The matrix $\mathbf{\Lambda}$ is first converted into a grayscale image - with the darker regions being defined as the points of higher intensity using $-\mathbf{\Lambda}$. The grayscale image is mapped such that the values in $-\mathbf{\Lambda}$ that are less than or equal to $\min(-\mathbf{\Lambda})*0.5$ are set to 0 (black), and values greater than or equal to $\max(-\mathbf{\Lambda})$ are set to 1 (white). This compresses the dynamic range of the directional spectrum ensuring that only the main discrete reflections in the time-frame are detected.

To allow reflections with overlapping spatial regions to be detected as individual events, image processing and segmentation algorithms from MATLAB's *Image Processing Toolbox* are used to separate overlapping regions in the directional spectrum. Firstly, a binary image is computed for the grayscale image extracting the darkest

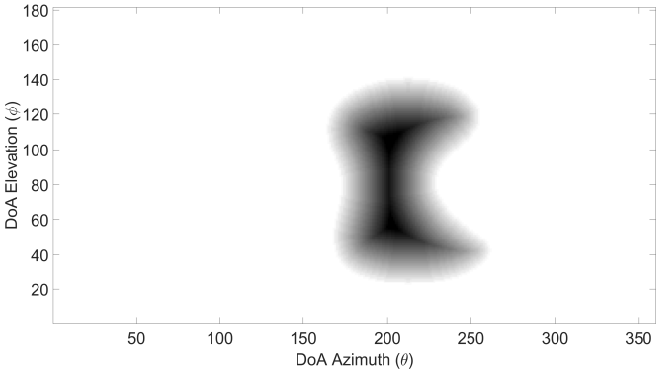

FIG. 2. An example of the binary mask of the directional spectrum (as seen in Figure 1) after having an extended minima mask applied. As can be seen the spatial region for the two arriving reflections overlap.

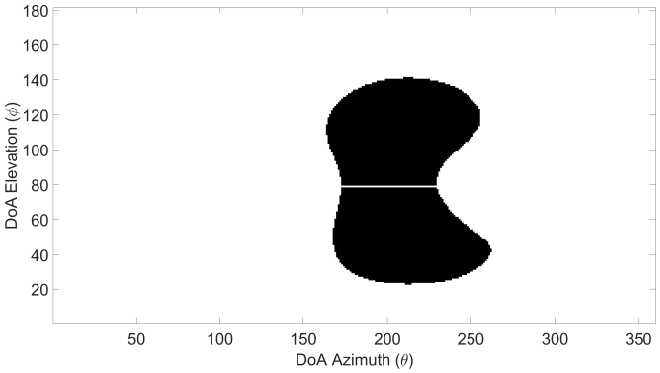

FIG. 3. An example of the resulting directional regions after being masked by the output of the watershed algorithm. The overlapping regions as seen in Figure 2 are now separated by a white line.

regions of the directional spectrum as the indexes with a color value less than ϵ_{msk}. To optimise the segmentation procedure, an extended-minima mask is applied to the binary image producing more defined regions, as shown in Figure 2. This extended-minima mask is created from the distance transform (*distanceTransform = -bwdist(~binaryImage)*[22]) of the binary image using the *imextendedmin* function[23], and applied to the binary matrix using the *imimposemin* function[24]. The *watershed* function[25] with default parameters is then used to create a label matrix containing positive valued integers for each region, with the regions separated by zero valued indices. This label matrix is applied to the binary image, by setting the indexes in the matrix where the watershed algorithm outputs a 0 to 0, as shown in Figure 3.

The dark regions in the masked directional spectrum are then detected using MATLAB's *regionprops* function[26], called as *regionprops(watershedMaskedImage, directionalSpectrum, 'all')*. The convex hull for these detected regions, as shown in Figure 4, represent the arrival of discrete reflections. When considering a 2D image of an unwrapped sphere, the wrap around observed at azimuth angles at $\theta = 0°$ and $\theta = 359°$ is not represented. Therefore, a directional signal occupying this spatial region will exist as two separate dark areas on the directional spectrum. To this extent, if two regions, with the same elevation values, exist at $\theta = 0°$ and $\theta = 359°$, the regions are combined and considered to be the arrival of a single reflection. Furthermore, it is important to note that this approach will not be able to distinguish between two reflections arriving from a similar DoA and ToA, and will detect these as a single reflection.

As this is an overlapping iterative process, and each reflection occupies a range of samples in the SRIR, the same reflection can be present across multiple subsequent time-frames. Therefore, each detection is either a reflection that was detected in the previous time-frame, or a new reflection. To resolve this ambiguity, the spatial region for each detection within the current time-frame is compared to any detections in the previous time-frame, and if any spatial region in the current time-frame had an overlap of at least 80% with any in the previous time-frame, they are considered to have been produced by the same reflection. The value of 80% is chosen to try and prevent individual reflections arriving from similar directions being detected as the same reflection. All reflections are therefore considered unresolved until their spatial region is no longer present in a subsequent time frame, a time-frame is skipped, or the iterative process ends. Once a detected reflection has been resolved using this process, the spatial and temporal region for the reflection is known and can be used to estimate the ToA and DoA.

The DoA can be estimated from the spatial region within each time-frame for which a reflection is present. The DoA is computed by adding the directional spectrum across the reflection's time-frames, and taking the steered direction, within the reflection's spatial region, with the largest intensity as corresponding to the DoA of the reflection:

$$\Psi_{DoA} = \arg\max_{\Psi} \left(\sum_{i=1}^{i=I} \mathbf{\Lambda}_i(\Psi_r) \right) \tag{9}$$

where $\mathbf{\Lambda_i}$ is the directional spectrum matrix for the ith time-frame that the reflection is present, r defines the sub-array indices in Ψ that define the spatial region, I is the total number of time-frames over which the reflection is present, and $\arg\max(...)$ outputs the steered direction with highest intensity value.

The ToA for the reflection is then defined as the time index containing the maximum peak present in the time-frame of the RIR, starting at the first window that the reflection is present, and ending at the start of the subsequent window where the reflection is no longer present. To distinguish between multiple reflections in a single time-frame, the response of the microphone is steered in

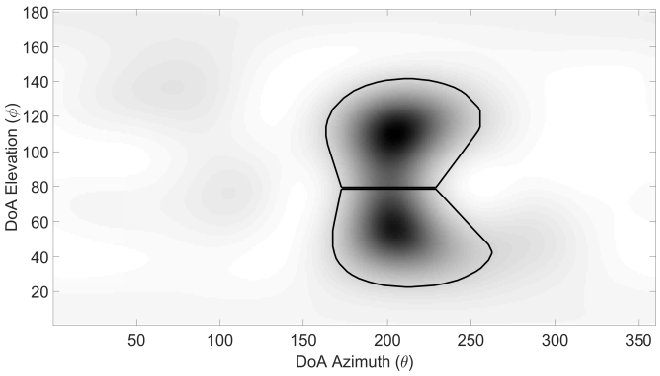

FIG. 4. An example of the detected regions (black contours) within the directional spectrum of Figure 1.

the direction of the DoA of the arriving reflection through spatial filtering with the spherical harmonic transform as,

$$ToA = \arg\max_{\tau}(|\mathbf{y}(\Psi_{DoA})\mathbf{H}^{T}(t-\tau_r)|) \tag{10}$$

where ToA is the time of arrival for the given reflection, τ_r is the time-frame that the reflection occupies in the SRIR, Ψ_{DoA} contains the azimuth and elevation DoA for the reflection, $\arg\max(...)$ returns the index where the maximum value of the expression is, and $|...|$ denotes absolute value.

B. The Acoustic Reflection Cartographer Method

Once the discrete reflections present in the SRIRs have been detected, the geometry of the room can be inferred. The proposed geometry inference method, The Acoustic Reflection Cartographer (E-ARC), has two processing stages: image-source reversion and geometry validation.

Before any processing is done, the detected reflections across all SRIRs are combined into one structure, and organised in descending ToA.

1. Image-source Reversion

For each candidate detection an estimated ToA and DoA value will be extracted from the SRIR. Assuming that the first arrival at the microphone array belongs to the direct sound and all subsequent detections are reflections, the source location (if not known *a priori*) and image-source locations can be defined using directional cosines, from[27] as,

$$\widetilde{\mathbf{s}}_i = \mathbf{m} + d_i \begin{bmatrix} \sin(\phi_i)\cos(\theta_i) \\ \sin(\phi_i)\sin(\theta_i) \\ \cos(\phi_i) \end{bmatrix} \tag{11}$$

where $\mathbf{m}$ is the $[x, y, z]$ coordinate for the receiver and d_i is the distance travelled by the ith detection, computed as ToA $* c$, where c is the speed of sound. To define the parameters for the candidate boundaries (4), the most-likely previous source for each image-source needs to be found, and substituted for $\mathbf{s}$ in (4).

For notation purposes in this section $\widetilde{\mathbf{s}}_i$ will be used to refer the current image-source being tested, $\widetilde{\mathbf{s}}_k$ is a candidate previous-source inferred from the kth reflection which has a ToA less than the reflection that produced $\widetilde{\mathbf{s}}_i$, and $\widetilde{\mathbf{s}}_j$ is the already detected previous-source for the candidate previous-source $\widetilde{\mathbf{s}}_k$.

When searching for the most-likely previous-sources it is important to consider that each image-source is either produced by a first-order reflection from a new or existing boundary, a higher-order reflection from an existing boundary, or a false-positive detection. Following the definition of the SRIR in Section III, it can be assumed that at least the first two detections after the direct sound from each SRIR, that are a consequence boundary more than 50 cm away from the source and receiver, are first-order. Furthermore, it is assumed that the first detection that can produce either the floor or ceiling for each source/receiver pair is first-order, and that the mean boundary position for these is assumed to be the floor and ceiling location. For subsequent reflections the assumption of first-order does not hold, as the first arriving second-order reflection will likely arrive before the last first-order[2]. Therefore, for subsequent reflections the most likely previous-source needs to be found, which will either be the source location or an image-source produced by a reflection with a ToA less than the reflection that produced the image-source being analysed. This search is constrained based on the aforementioned assumptions, which ideally will eliminate the majority of false-positive detections made by the reflection detection method.

The first consideration in the process is to ascertain whether the image-source is as a result of a reflection from a known boundary. This is tested for the source and all image-sources ($\widetilde{\mathbf{s}}_k$) that are defined by reflections with a ToA less than that of the reflection that produced $\widetilde{\mathbf{s}}_i$ as,

$$\text{previousSource} = \widetilde{\mathbf{s}}_k, \text{ if } ||(\widetilde{\mathbf{s}}_k + 2\,\langle \widetilde{\mathbf{b}}_l - \widetilde{\mathbf{s}}_k, \hat{\mathbf{n}}_l\rangle \hat{\mathbf{n}}_l) - \widetilde{\mathbf{s}}_i|| \leq \epsilon_{\widetilde{\mathbf{s}}} \tag{12}$$

where $\widetilde{\mathbf{s}}_k$ is the image-source for the kth reflection, $l = 1 : L$ is the number of inferred boundaries defined by first-order reflections, $\langle .,. \rangle$ denotes dot product, and $\epsilon_{\widetilde{\mathbf{s}}}$ is an empirically defined threshold value chosen to allow for inaccuracies in ToA and DoA estimation. If any of these image-sources tested produce an image-source location close to the actual image-source ($\widetilde{\mathbf{s}}_i$) it is assumed to be the most-likely previous-source.

If no existing boundaries defined by a first-order reflection are attributable to $\widetilde{\mathbf{s}}_i$, then a new boundary is defined. As with the previous work in the literature[1,2,5] an image-source that cannot be defined using existing boundaries is assumed to be first-order. However, contrary to these works a set of constraints are imposed to remove image-sources that are as a result of false-positive detections, these constraints are (Figure 5),

- The difference in propagation distance Δl between the image-source-to-receiver path and source-to-boundary-to-receiver path should be within a defined threshold such that $\Delta l \leq \epsilon_l$, where ϵ_l is the threshold

- The inferred boundary is perpendicular to the floor and ceiling, defined using the z-axis coefficient for the boundary's normal $\widetilde{\mathbf{n}}_z$, which should be $\widetilde{\mathbf{n}}_z = 0$, constrained as $\widetilde{\mathbf{n}}_z \leq \epsilon_{\widetilde{\mathbf{n}}}$, where $\epsilon_{\widetilde{\mathbf{n}}}$ is the threshold value.

- The inferred boundary is at least 50 cm away from the source and receiver, as defined by the minimum source-to-boundary and receiver-to-boundary distances.

- The specular reflection produced by the path from source-to-boundary should have x and y directional cosines close to that of the actual reflection path from image-source-to-receiver, such that $\Delta\angle \leq \epsilon_\angle$ where $\epsilon_\angle$ is the threshold value used and $\Delta\angle$ is calculated as $||[\widetilde{\alpha}, \widetilde{\beta}] - [\alpha, \beta]||$ and $\widetilde{\alpha}$, $\widetilde{\beta}$ are calculated, using ray-tracing[27], as,

$$\widetilde{\alpha} = \widetilde{\alpha}_{prev} - 2cos(\upsilon)\mu \tag{13}$$
$$\widetilde{\beta} = \widetilde{\beta}_{prev} - 2cos(\upsilon)\eta \tag{14}$$

  where $\widetilde{\alpha}$ and $\widetilde{\beta}$ are the directional cosines along the x and y axes respectively, α_{prev} and β_{prev} are the directional cosines computed for a line going from the previous-source to the point where the line from image-source-to-receiver intersects the boundary $\mathbf{b}_r$, υ is the angle of incidence, and μ and η are the directional cosines of the normal vector of the plane along the x and y axes.

The implication of the assumption of first-order is that any higher-order reflections defined as a first-order reflection will produce a boundary distant from the desired boundary location, and therefore, a geometry validation process is required to refine the inferred boundaries. Furthermore, second-order reflections that are produced by interactions between perpendicular boundaries will produce an angled boundary that will impact the inferred shape of the room. Therefore, attempts are made to find the correct previous-source for image-sources produced by these perpendicular reflections.

Reflections produced by interactions between perpendicular boundaries are searched for by exploiting the properties of the image-source method. Given that an image-source is generated by mirroring its previous-source perpendicularly across a boundary, for the case of a reflection between perpendicular boundaries, the relationship between the image-source, previous-source, and the previous-source of the previous-source, can be expressed as a rotation of these image-sources around a point in space (Figure 6). From this relationship, this

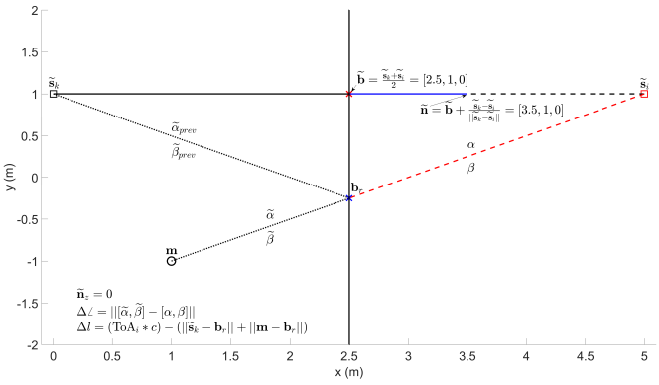

FIG. 5. Example of the relationship between the previous-source $\widetilde{\mathbf{s}}_k$, image-source $\widetilde{\mathbf{s}}_i$, and a reflective boundary. The dotted line denotes the reflection path computed for a specular reflection using (13) and the red dashed line is the path from image-source to receiver $\mathbf{m}$ (color online).

point of rotation must be equidistant from these three image-source locations in the reflection path. Therefore from (4) this point of rotation $\mathbf{p_r}$ can be expressed as,

$$\mathbf{p_r} = \frac{\widetilde{\mathbf{s}}_i + \widetilde{\mathbf{s}}_j}{2} \tag{15}$$

where $\widetilde{\mathbf{s}}_i$ is the image-source being analysed and $\widetilde{\mathbf{s}}_j$ is the previous-source of the previous-source. The image-source produced for a reflection between perpendicular boundaries can therefore be detected if the image-source and previous-source are equidistant from this point of rotation as,

$$\text{previousSoure} = \widetilde{\mathbf{s}}_k, \text{ if } |\,||\widetilde{\mathbf{s}}_i - \mathbf{p_r}|| - ||\widetilde{\mathbf{s}}_k - \mathbf{p_r}||\,| \leq \epsilon_o \tag{16}$$

If more than one previous-source can be defined using this relationship then the previous-source with the smallest error in reflection path is used as,

$$min(\Delta l + \Delta\angle + \widetilde{\mathbf{n}}_z) \tag{17}$$

In the case that none of these steps produce a valid candidate previous-source, the image-source in question is assumed to be as a result of a false-positive detection made by the reflection detection method. An overview of this process can be seen in Algorithm 1.

2. Geometry Validation

From Figure 7 (a), it can be seen that there are three types of potentially erroneous boundary detections. (I) Boundaries positioned on the corners of the desired geometry as a result of a correct previous-source not being detected for a second-order reflection between perpendicular boundaries. (II) Boundaries positioned immediately after another boundary, which are likely to be a product of either noise, or a single reflection being detected as multiple separate arrivals. (III) Boundaries positioned

Generate image-sources
for i = *number of reflection : -1 : 1* **do**
  if $\widetilde{\mathbf{s}}_i$ *is within 1 meter of source or receiver* **then**
    remove $\widetilde{\mathbf{s}}_i$ as it cannot produce a valid boundary
    continue
  end
  if *number of detections* < *2* **then**
    if ***norm***$(((\widetilde{\mathbf{s}}_i + \mathbf{s})/2) - \mathbf{s}) > 0.5$ ***and*** ***norm***$(((\widetilde{\mathbf{s}}_i + \mathbf{s})/2) - \mathbf{m}) > 0.5$ **then**
      previousSource$_i$ = $\mathbf{s}$
      number of detections = number of detections + 1
      continue
    end
  end
  for $l = 1 : L$ **do**
    for $k = i + 1$: *number of reflections* **do**
      if *norm*$(\widetilde{\mathbf{s}}_k + 2 *$ *dot*$(\widetilde{\mathbf{b}}_\mathbf{l} - \widetilde{\mathbf{s}}_k, \widetilde{\mathbf{n}}_\mathbf{l}\text{-}\widetilde{\mathbf{s}}_i) < \epsilon_s$ **then**
        previousSource$_i$ = $\widetilde{\mathbf{s}}_\mathbf{k}$
        number of detections = number of detections + 1
        continue
      end
    end
  end
  Define new boundary using source as previous-source
  if $\Delta l < \epsilon_l$ ***and*** $\Delta\angle < \epsilon_\angle$ ***and*** $\Delta\widetilde{\mathbf{n}} < \epsilon_{\widetilde{\mathbf{n}}}$ **then**
    possiblePreviousSource = $\mathbf{s}$;
    if *bounday is the first that can define the floor or ceiling* **then**
      previousSource$_i$ = $\mathbf{s}$
      number of detections = number of detections + 1
      continue
    end
  end
  store = 1
  for $k = i + 1$: *number of reflections* **do**
    if $\widetilde{\mathbf{s}}_\mathbf{k}$ *and* ($\widetilde{\mathbf{s}}_\mathbf{i}$ *have a difference in distance* $\leq \epsilon_o$ *to* $\mathbf{p_r}$ **then**
      possiblePreviousSource(store, :) = $\widetilde{\mathbf{s}}_\mathbf{k}$
      store = store + 1
    end
  end
  if ***length***(*possiblePreviousSource*) == *0* **then**
    remove $\widetilde{\mathbf{s}}_\mathbf{i}$
    continue
  end
  if ***length***(*possiblePreviousSource*) > *1* **then**
    [, minIndex] = **min**$(\Delta l + \Delta\angle + \widetilde{\mathbf{n}}_z)$
    previousSource(ii,:) = possiblePreviousSource(minIndex, :)
    number of detections = number of detections + 1
    else
      previousSource$_i$ = possiblePreviousSource
      number of detections = number of detections + 1
    end
  end
end

Algorithm 1: Pseudocode for image-source reversion process considering a single source and receiver.

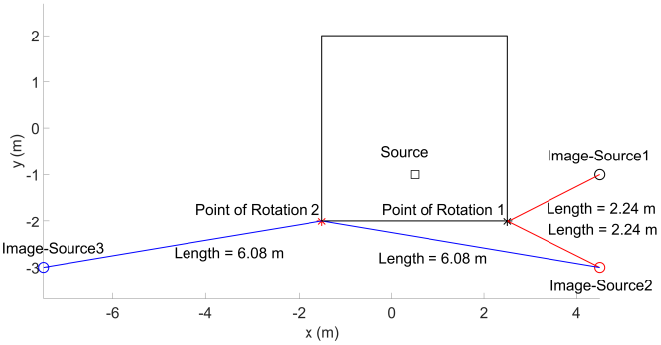

FIG. 6. Diagram showing the rotational relationship between the image-source and its previous source, in this case Image-Source2 with Image-Source1 and Image-Source3 with Image-Source2 (color online). Image-Source1 is produced by mirroring the source in the boundary on the right side of the square. Image-Source2 is produced by mirroring Image-Source1 in the lower boundary, and Image-Source3 is produced by mirroring Image-Source2 in the left boundary. Point of Rotation 1 is the mid-point between Image-Source2 and the Source location, and Point of Rotation 2 is the mid-point between Image-Source3 and Image-Source1.

far outside of the desired geometry, which are as a result of higher-order reflections being defined as first-order. The latter two of these potentially erroneous boundary conditions will be considered here, as they will have the largest impact on the accuracy of the geometry inference process.

Ahead of the next step, boundaries that are coincident are removed until only one remains, reducing the number of boundaries to be tested. Two boundaries are defined as being coincident if the boundary normals $\widetilde{\mathbf{n}}_1$ and $\widetilde{\mathbf{n}}_2$ are parallel and the inferred point on the boundaries $\widetilde{\mathbf{b}}_1$ and $\widetilde{\mathbf{b}}_2$, where $_1$ and $_2$ denote different boundaries, exists on both boundaries[35], such that,

$$||\widetilde{\mathbf{n}}_1 \times \widetilde{\mathbf{n}}_2|| \leq \epsilon_{par} \tag{18}$$

$$\text{and } | < \widetilde{\mathbf{n}}_1, \widetilde{\mathbf{b}}_1 - \widetilde{\mathbf{b}}_2 > | \leq \epsilon_{point} \tag{19}$$

where ϵ_{par} and ϵ_{point} are empirically defined threshold values to account for small variations in boundary position as a result of ToA and DoA errors. An additional constraint is required to account for non-convex-shaped rooms, where boundaries that are mathematically coincident could define two separate boundaries. To then remove the aforementioned inferred boundaries that are positioned outside of the desired geometry of the room, a three step geometry validation process is proposed. These three steps are as follows, **reflection path validation**, **line-of-sight boundary validation**, and **closed geometry validation** (Algorithm 2). Each of these steps are performed on an approximate estimation of the inferred room's geometry, based on the nearest intersection points[35] between non-parallel inferred boundaries.

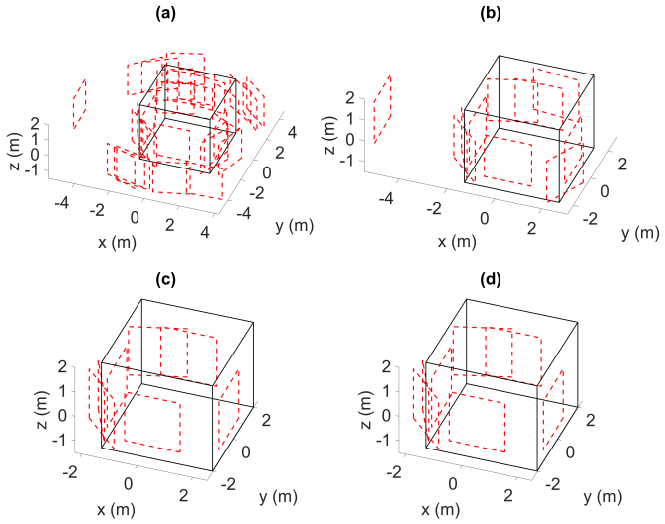

FIG. 7. Example showing the remaining inferred boundaries after each step of the geometry validation process. Figure (a) shows the unconstrained boundaries before any processing is performed, Figure (b) shows the remaining unconstrained boundaries after the Reflection Path Validation step, Figure (c) shows the remaining boundaries after the line-of-sight test, and Figure (d) shows the remaining unconstrained boundaries after the closed geometry test (color online).

Step 1: Reflection Path Validation

The first step is to check if the reflection path from the image-source-to-receiver is obstructed by additional boundaries that are closer to the receiver than the boundary inferred by this image-source. This step will remove the majority of boundaries positioned outside of the desired room's geometry, as the reflection path for these boundaries will be occluded by the candidate boundaries that define the geometry of the room, as seen in Figure 7 (a). This step is performed by defining a line from the image-source that produced the boundary being tested to the receiver, and computing the intersection[36] between the line and every other boundary. If any other boundary occlude the reflection path from the ith boundary to the receiver, the ith boundary is removed. Once all boundaries have been tested, the shape of the room is inferred from the remaining boundaries and the process is repeated until no further boundaries are removed. An example of the resulting inferred geometry after this step can be seen in Figure 7 (b), where only one external boundary has not been removed.

Step 2: Line-of-Sight Boundary Validation

While the majority of incorrect boundaries have now been removed, there are still non-valid boundaries that exist as a result of the path from image-source to receiver not passing through the boundary the image-source produces. Therefore, a line-of-sight test is performed to ensure all inferred boundaries are visible to at least one receiver position. Any boundaries that are not in line-of-sight of the receiver could not have produced a reflection that arrives at the receiver. To test line-of-sight, a set of rays are defined with $0 \leq \theta \leq 359$ and $\phi = 90$ using (11), with an arbitrary value for d_i. The first boundary that intersects[36] with each of these rays is considered valid. An example of the resulting inferred boundaries after this step has been performed can be seen in Figure 7 (c), where the remaining additional boundary from Figure 7 (b) has been removed.

Step 3: Closed Geometry Test

These first two steps will have refined candidate boundaries to that of the desired room for the majority of cases. The final step is to ensure that the inferred geometry of the room produces a closed shape. As with the previous two stages the geometry of the room is first inferred, then any constrained boundaries that do not intersect with two adjacent boundaries, one on each side, are removed. An example of the resulting inferred boundaries after this step has been performed can be seen in Figure 7 (d), in this case the inferred room had produced a closed geometry.

IV. TESTING

To test the proposed method, three scenarios are presented.

Scenario One consists of three sets of simulations, one for a convex Cuboid-Shaped room, a convex Octagonal-Shaped Room, a non-convex L-Shaped Room, and a non-convex T-Shaped Room. The Cuboid, Octagonal, and L-shaped room consist of two measurement positions, and the T-Shaped room three, simulated using CATT-Acoustic[28]. The SRIRs were simulated using omnidirectional sources at three different locations. The simulations were run with 10,000,000 rays ensuring sufficient excitation of the entire room, with diffuse reflections turned off. The boundaries are defined as having the absorption properties of a wooden surface, using the WOOD30 material in CATT-Acoustic[28]. The resulting SRIRs are rendered out as third-order spherical harmonic domain signals. The geometry and source and receiver positions can be seen in Figure 8.

Scenario Two consists of two L-shaped rooms, with volumes 320 m^3 and 360 m^3, simulated in CATT-Acoustic using the same parameters as the L-Shaped room in Scenario One. These rooms are simulated using a single receiver positioned in line-of-sight of every boundary, and 15 randomly selected source positions in each segment of the room. From these two sets of 15 source positions for each L-shaped room, a selection of 33 source combinations that ensure a first-order reflection from each boundary, are used to test the proposed method. This example tests the variability of the performance of the method, quantifying any difference in estimation accuracy between the two rooms.

Scenario Three consists of two RIR measurements captured using the EigenMike EM32, a 32-channel

```
while changesMade ~= 0 do
    Infer geometry using boundary-boundary
     intersections.
    changeHappened = 0
    Step 1: Check reflection path for
     multiple boundary intersections
    for i = 1 : numberOfBoundaries do
        for k = 1 : numberOfBoundaries do
            if boundary k intersects line going
             from point of incidence on boundary
             i and the receiver then
                remove boundary i
                changeHappened = 1;
            end
        end
    end
    if changeHappened == 0 then
        changesMade = 0
    end
end
Step 2: line-of-sight test
for θ = 1 : 359 do
    Define ray in azimuth direction θ
    for i = 1 : numberOfBoundaries do
        if ray intersects boundary i then
            boundaryIsValid(i) = 1
        end
    end
    Remove boundaries where boundaryIsValid
     == 0
end
Infer geometry using boundary-boundary
 intersections.
Step 3: Closed Geometry test
for ii = 1 : numberOfBoundaries do
    Compute the distance between boundary ii
     and adjacent boundaries
    if boundaries do not connect and distance
     between boundaries is < 0.1 then
        remove boundary ii
    end
end
Infer geometry using boundary-boundary
 intersection points.
```

Algorithm 2: Pseudocode for Geometric Validation process

spherical microphone array, and a Genelec 8030 loudspeaker. The sound source used to capture the response of the room was an exponential sine-sweep[30] 20 seconds in length with a frequency range of 100-20 kHz, using the inverse-filter of the original sine-sweep to produce the SRIR. To better approximate an omnidirectional source, an average of the RIRs measured at four speaker orientations (0°, 90°, 180°, and 270°) is taken[31]. The final SRIRs are then normalised to have a maximum sample value of ±1, and converted to third-order spherical harmonic domain signals using MH Acoustics' EigenStudio[32]. The measurement room is cuboid-shaped with dimensions 10.35 m×13.29 m×4.19 m, and has a number of non-removable, adjustable, floor length curtains. As it was not possible to remove these curtains, they were positioned, as much as is possible, to limit their impact on the obtained SRIRs. Hence they were arranged in corners of the room, across windows, and, where possible, to cover features on the walls such as electrical outputs, as well as the computer and interface used for the measurements. While it is accepted that this is non-ideal, and could have some impact on the results, every effort has been made to minimize their potential influence on the measurements obtained, and ensure that the main reflective boundaries are exposed and clear from other possibly confounding features. Furthermore, the ceiling was covered in large metal piping connected to extractor fans and a layer of metal railing approximately 1 m from the ceiling. The noise floor in the room is measured as 60.2 dBA and the room's temperature was 24.4°C, and hence the speed of sound is estimated as 346.97 m/s. The room's geometry, loudspeaker and receiver positions, and an image of the room can be seen in Figure 9.

The total measurements needed in these scenarios are defined by the number needed to ensure first-order reflections for each boundary are captured. In practice any number of SRIRs can be used. The SRIRs being analysed are truncated to 3000 and 2000 samples in length for the measured and simulated examples respectively. The truncation lengths are chosen to allow sufficient time for the main reflections to arrive while reducing the number of higher-order reflections.

Three error metrics are used to analyse the accuracy of the inferred boundaries. (I) ΔPosition - the distance between desired and inferred boundaries[1,2,7,10] measured at 10 cm intervals along the length of the target boundary and the RMS error computed over these intervals. (II) Dihedral Angle[7,8,10] - the angle between desired and inferred boundaries. Dihedral angle is averaged over inferred boundaries if more than one exists. (III) ΔLength[2] - difference between desired and inferred boundary length.

The threshold values used were derived empirically through examination of results obtained for different cases, and chosen so all first-order reflections are detected, while reducing the number of inaccurately inferred boundaries due to noise: $\epsilon_a = 0.01$, $\epsilon_d = 30\%$, $\epsilon_{msk} = 0.1$, $\epsilon_{\tilde{\mathbf{s}}} = 30$ cm, ϵ_{par}, 0.1, $\epsilon_{point} = 0.1$, and $\epsilon_O = 15$ cm.

The SRIRs used in testing are shifted in time to ensure that the ToA of the direct sound matches that expected given the source-receiver distances and speed of sound, removing latency introduced by the measurement system. Furthermore, the DoA is shifted to ensure that $\theta = 0°$ is aligned with the positive going x axis.

V. RESULTS

A. Scenario One

When analysing simulated SRIRs, further consideration of the signal properties, and therefore choice of beamformer is required. When simulating SRIRs the arriving reflections are highly correlated as a result of each reflection being treated as a perfect Dirac, which in real-world measurement conditions is not the case. Therefore, if there are multiple reflections, the directional spectrum

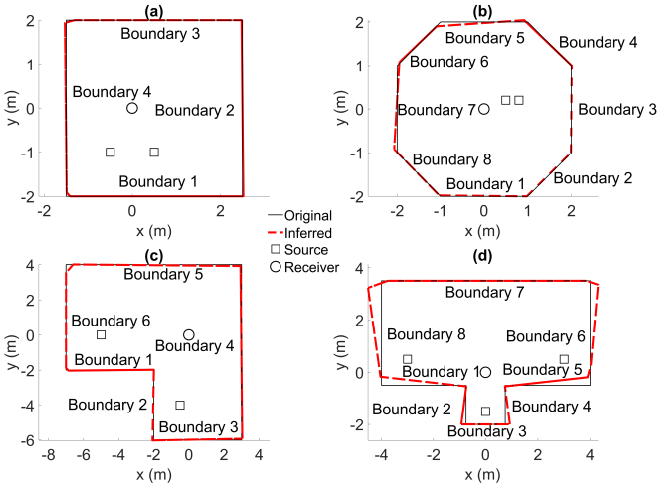

FIG. 8. Inferred geometry (dashed line) and desired geometry (solid line) for Scenario One Cuboid-Shaped Room (a), Octagonal-Shaped Room (b), L-Shaped Room (c), and T-Shaped Room (d) (color online).

Room	ΔPosition	Dihedral Angle	ΔLength
Cuboid	4.63 cm	8.59°	6.32 cm
Octagonal	2.69 cm	2.01°	9.01 cm
L-Shaped	4.69 cm	14.02°	25.92 cm
T-Shaped	16.45 cm	8.03°	42.45 cm

TABLE I. Results for Scenario One - simulated data for the Cuboid, Octagonal, L-shaped, and T-shaped rooms: presenting the RMS difference in position (ΔPosition), dihedral angle, and difference in boundary length (Δlength).

produced by the MVDR beamformer will have a lower range between the directional response of the reflections and the directional response of the residual signal, as the signals covariance matrix is rank-deficient[33]. This reduction in range can lead to first-order reflections being missed. Therefore, when analysing the simulated impulse responses the plane wave decomposition[14,34] beamformer is used instead.

The results in Table I and Figure 8 show that for the Cuboid, Octagonal-, and L-Shaped Room, the shape of the room has been inferred, with RMS Δ Positions of 4.63 cm, 2.69 cm, and 4.69 cm respectively. However, the boundaries of the T-Shaped room are angled, resulting in an increase in RMS boundary position errors. For the Cuboid- and L-Shaped room the large dihedral angles are a result of boundaries being inferred in the corners of the room due to incorrectly assigned previous-sources for second-order reflections. However, the dihedral angles for Octagonal- and T-Shaped rooms are attributable to inferred boundaries being slightly angled, as a result of ToA and/or DoA estimation errors for the reflections.

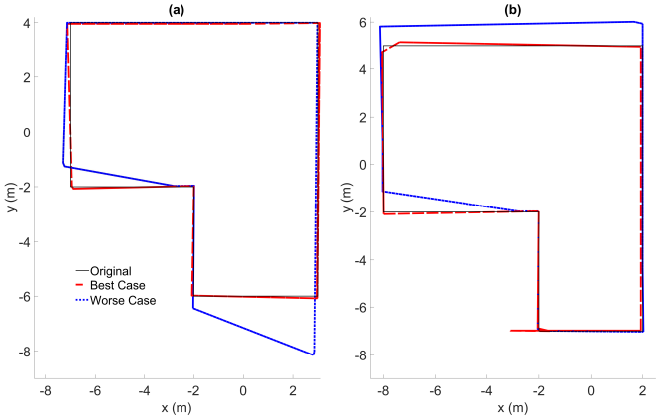

FIG. 9. Desired geometry (solid line), best case inferred geometry (red dashed line), and worst case inferred geometry (blue dotted line) for Scenario Two L-Shaped Room One (a) and Scenario Two L-Shaped Room Two (b) (color online).

B. Scenario Two

The results in Table II are presented as average boundary errors across the test cases. To perform statistical analysis of the data the non-parametric Kruskal-Wallis test, *kruskalwallis*[38], is used, and reported as (χ^2 = , p = , degrees of freedom =). Furthermore, the bootstrap process[37] is used to compute the 95% confidence interval for the mean values using *bootstrapci*[39].

There is a 7.41 cm difference between the mean boundary positional error across the two L-shaped rooms, with the second, larger, L-Shaped room having larger boundary positional errors. This increase in boundary position error is a consequence of 11 cases for the second L-shaped room having an additional angled boundary being inferred, compared to five for the first. The minimum and maximum mean error for the measurement sets in L-shaped Room One are ΔPosition = [3.95 cm, 35.58 cm], Dihedral Angle = [2.24°, 11.22°], and ΔLength = [6.48 cm, 110.98 cm], and for the second L-Shaped Room, ΔPosition = [4.22 cm, 32.81 cm], Dihedral Angle = [1.05°, 10.30°], and ΔLength = [8.40 cm, 85.95 cm]. Comparing the variation in measurement accuracy between the two L-shaped rooms shows no significant difference for the ΔPosition ($\chi^2 = 0.0005$, $p = 0.98$, degrees of freedom = 395) and ΔLength ($\chi^2 = 0.35$, $p = 0.55$, degrees of freedom = 395). However, there is a significant difference for the dihedral angle ($\chi^2 = 10.25$, $p = 0.0014$, degrees of freedom = 395), as a result of the additional boundaries inferred in the corners of the room for L-Shaped Room One. This suggests that while there are differences in the mean values between these two examples, the variability in performance between the two sets are comparable. The best and worse cases for each L-Shaped room can be seen in Figure 9.

Room	ΔPosition	Dihedral Angle	ΔLength
One	11.50±0.1 cm	7.28°±0.048°	36.84±0.31 cm
Two	18.91±0.17 cm	3.69°±0.027°	5.63±0.36 cm

TABLE II. Average error metrics, difference in position (ΔPosition), dihedral angle, and difference in boundary length (Δlength), for Scenario Two L-Shaped Rooms One and Two.

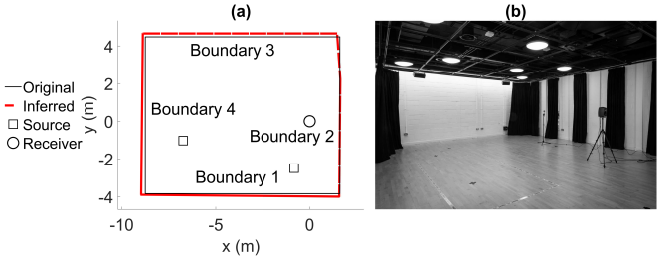

FIG. 10. Figure (a): Inferred geometry (dashed line) and desired geometry (solid line) for Scenario Two (color online). Figure (b): image of room setup.

C. Scenario Three

As can be seen in Figure 10 and the results in Table III, the general shape of the room has been inferred with small dihedral angle values between the original and inferred boundaries. However, there are some inaccuracies in the boundary positions, and therefore the lengths of surrounding boundaries. These inaccuracies are likely due to either imperfect specular reflections, under or over estimation of the ToA for reflections in the measured impulse responses, or any inaccuracy in the estimated DoA of the reflections. These lead to incorrect estimation of the image-sources desired position, which affects both the positioning of the boundary it infers, and subsequent boundaries that are defined using this image-source.

Boundary	ΔPosition	Dihedral Angle	ΔLength
1	16.16 cm	0.54°	18.45 cm
2	4.02 cm	2.28°	34.63 cm
3	25.46 cm	0°	1.4 cm
4	13 cm	0.60°	24.05 cm
Floor	14.5 cm	N/A	N/A
Ceiling	10.60 cm	N/A	N/A
RMS	15.37 cm	1.21°	23.02 cm

TABLE III. Results for Scenario Three, real-world measurements, presenting the difference in position (ΔPosition), dihedral angle, and difference in boundary length (Δlength)

VI. DISCUSSION

The results presented show comparable RMS boundary position estimation error to those in[1,5,8,10], which reported error values between 0.063 cm to 29.38 cm when considering only cuboid-shaped rooms. The results showed that the T-Shaped Room had the largest estimation error, as a result of the inferred boundaries being angled. This could be as a result of the room's complexity, or the requirement for more measurement positions, which increases the chance of erroneous boundaries being estimated. Results presented in Scenario Two showed that the proposed method displayed statistically similar performance across two set of 33 measurements for two L-Shaped rooms.

VII. CONCLUSIONS

This paper has presented a new method reflection detection and geometry inference, without needing large numbers of measurement locations or an assumption of convexity of the measured room. This is achieved by exploiting directional information contained within spatial room impulse responses measured using a spherical microphone array. The method is tested using simulated SRIRs for two convex- and four non-convex-shaped rooms, and real-world measurements in a cuboid room. The results showed that the inferred room's geometry is close to that of the desired, with generally low RMS boundary distance errors and errors in dihedral angle generally as a result of angled boundaries in the corners of the room. The RMS boundary position errors are comparable to prior work[1,5,8,10] with a maximum difference in RMS error of 16.38 cm, using at most three measurement positions, compared to 6-64 used in this prior work. In addition to this, results presented in Scenario Two showed that the variance in performance of the proposed method is statistically similar across the two cases presented. Further investigation into non-convex geometry inference could explore alternative means of retracting higher-order reflection paths to improve robustness, and should consider real-world measurements of non-convex shaped rooms. Furthermore, subspace-beamformers could be explored to try and improve the robustness of the reflection detection process to interfering noise.

ACKNOWLEDGMENTS

Funding was provided by a UK Engineering and Physical Sciences Research Council (EPSRC) Doctoral Training Award.

[1]L. Remaggi, P. J. B. Jackson, P. Coleman, and W. Wang, "Acoustic Reflector Localization : Novel Image Source Reversion and Direct Localization Methods," IEEE/ACM Transactions on Audio, Speech and Language Processing **25**(2), 296–309 (2017) doi: 10.1109/TASLP.2016.2633802.

[2]I. Dokmanic, R. Parhizkar, A. Walther, Y. M. Lu, and M. Vetterli, "Acoustic echoes reveal room shape," in *Proceedings of the*

National Academy of Sciences (2013), Vol. 110, pp. 12186–12191, doi: 10.1073/pnas.1221464110.

[3]D. Arteaga, D. Gracia-Garzon, T. Mateos, and J. Usher, "Scene inference from audio," in *AES Convention 134*, Rome, Italy (2013), http://www.aes.org/e-lib/browse.cfm?elib=16794.

[4]F. Ribeiro, D. Florêncio, D. Ba, and C. Zhang, "Geometrically constrained room modeling with compact microphone arrays," IEEE Transactions on Audio, Speech and Language Processing **20**(5), 1449–1460 (2012) doi: 10.1109/TASL.2011.2180897.

[5]S. Tervo and T. Tossavainen, "3D Room Geometry Estimation from Measured Impulse Responses," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, Kyoto, Japan (2012), pp. 513–516, doi: 10.1109/ICASSP.2012.6287929.

[6]M. Kuster, D. de Vries, E. M. Hulsebos, and A. Gisolf, "Acoustic imaging in enclosed spaces: Analysis of room geometry modifications on the impulse response," The Journal of the Acoustical Society of America **116**, 2126–2137 (2004) doi: 10.1121/1.1785591.

[7]E. Nastasia, F. Antonacci, A. Sarti, and S. Tubaro, "Localization of planar acoustic reflectors through emission of controlled stimuli," European Signal Processing Conference 156–160 (2011) https://ieeexplore.ieee.org/document/7074175.

[8]J. Filos, A. Canclini, F. Antonacci, A. Sarti, P. A. Naylor, and P. Milano, "Localization of Planar Acoustic Reflectors from the Combination of Linear Estimates," 2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO) (Eusipco), 1019–1023 (2012) https://ieeexplore.ieee.org/document/6334299.

[9]L. Zamaninezhad, P. Annibale, and R. Rabenstein, "Localization of environmental reflectors from a single measured transfer function," ISCCSP 2014 - 2014 6th International Symposium on Communications, Control and Signal Processing, Proceedings 157–160 (2014) doi: 10.1109/ISCCSP.2014.6877839.

[10]Y. E. Baba, A. Walther, and E. A. Habets, "3D room geometry inference based on room impulse response stacks," IEEE/ACM Transactions on Audio Speech and Language Processing **26**(5), 857–872 (2018) doi: 10.1109/TASLP.2017.2784298.

[11]P. A. Naylor, A. Kounoudes, J. Gudnason, and M. Brookes, "Estimation of glottal closure instants in voiced speech using the DYPSA algorithm," IEEE Transactions on Audio, Speech and Language Processing **15**(1), 34–43 (2007) doi: 10.1109/TASL.2006.876878.

[12]D. P. Jarrett, E. A. Habets, and P. A. Naylor, *Theory and Applications of Spherical Microphone Array Processing* (Springer International Publishing, 2017).

[13]A. Politis, "getSH," https://github.com/polarch/Spherical-Harmonic-Transform/blob/master/getSH.m (2015).

[14]A. Politis, "Spherical Array Porcessing Toolbox," (2016), https://github.com/polarch/Spherical-Array-Processing.

[15]H. Sun,"Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing," The Journal of the Acoustical Society of America **131**(4), 2828–2840 (2012) https://asa.scitation.org/doi/abs/10.1121/1.3688476?journalCode=jas.

[16]D. P. Jarrett, E. A. Habets, and P. A. Naylor, "Spherical harmonic domain noise reduction using an MVDR beamformer and DOA-based second-order statistics estimation," ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings 654–658 (2013) doi: 10.1109/ICASSP.2013.6637729.

[17]BSI Standard ISO 3382-1 and British Standards Institution, "Acoustics - Measurements of room acoustic parameters Part 1: Performance Spaces (ISO 3382-1:2009)," (2009).

[18]B. Rafaely, B. Weiss, and E. Bachmat, "Spatial aliasing in spherical microphone arrays," IEEE Transactions on Signal Processing **55**(3), 1003–1010 (2007) doi: 10.1109/TSP.2006.888896.

[19]MH Acoustics, "Eigenbeam Datasheet," (2016), https://mhacoustics.com/sites/default/files/EigenbeamDatasheet_R01A.pdf.

[20]N. Epain and C. T. Jin, "Spherical Harmonic Signal Covariance and Sound Field Diffuseness," **24**(10), 1796–1807 (2016) doi: 10.1109/TASLP.2016.2585862.

[21]J. Capon, "High-resolution frequency-wavenumber spectrum analysis," Adaptive Antennas for Wireless Communications **57**(8), 1408–1418 (1969) doi: 10.1109/PROC.1969.7278.

[22]MathWorks, "bwdist," (2017), https://uk.mathworks.com/help/images/ref/bwdist.html.

[23]MathWorks, "imextendedmin," (2017), https://uk.mathworks.com/help/images/ref/imextendedmin.html.

[24]MathWorks, "imimposemin," (2017), https://uk.mathworks.com/help/images/ref/imimposemin.html.

[25]MathWorks, "Watershed," (2017), https://uk.mathworks.com/help/images/ref/watershed.html.

[26]MathWorks, "regionprops," (2017), https://uk.mathworks.com/help/images/ref/regionprops.html.

[27]A. Krokstad, S. Strøm, and S. Sørsdal, "Calculating the Acoustical Room Response by The Use of Ray Tracing Technique," Journal of Sound and Vibration **8**, 118–125 doi: 10.1016/0022-460X(68)90198-3.

[28]CATT, "CATT-Acoustic," (2016), http://www.catt.se/.

[29]MathWorks, "imclose," (2018), https://www.mathworks.com/help/images/ref/imclose.html.

[30]A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," Proc. AES 108th conv, Paris, France (2000) http://www.aes.org/e-lib/browse.cfm?elib=10211.

[31]F. Stevens and D. Murphy, "Spatial impulse response measurement in an urban environment," in *Presented at the 55th AES International Conference*, AES, Helsinki, Finland (2014), pp. 1–8, http://www.aes.org/e-lib/browse.cfm?elib=17355.

[32]MH Acoustics, "EigenStudio," (2013), https://mhacoustics.com/products.

[33]N. Huleihel and B. Rafaely, "Spherical array processing for acoustic analysis using room impulse responses and time-domain smoothing," The Journal of the Acoustical Society of America **133**(June 2013), 3995–4007 (2013) doi: 10.1121/1.4804314.

[34]D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, "3D Source localization in the spherical harmonic domain using a pseudointensity vector," in *Signal Processing Conference, 2010 18th European*, 5, IEEE, Aalborg, Denmark (2010), pp. 442–446, https://ieeexplore.ieee.org/document/7096575.

[35]E. Weisstein, "Plane-Plane Intersection," http://mathworld.wolfram.com/Plane-PlaneIntersection.html.

[36]E. Weisstein, "Line-Plane Intersection," http://mathworld.wolfram.com/Line-PlaneIntersection.html.

[37]B. Efron, "Bootstrap Methods: Another Look at the Jackknife," The Annals of Statistics **7**(1) (1979) https://projecteuclid.org/euclid.aos/1176344552.

[38]MathWorks, "kurkalwallis," (2006), https://uk.mathworks.com/help/stats/kruskalwallis.html.

[39]MathWorks, "bootstrapci," (2006), https://uk.mathworks.com/help/stats/bootci.html.